

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

(19) Japanese Patent Office (JP)

(12) Unexamined Patent Gazette (A)

(11) Unexamined Patent Application

No. 61-33570

(43) Published February 17, 1986

(51) Int. Cl.⁴ ID Symbol Internal File No.

G06F 15/20

A-711-5B

Request for Examination: not submitted

Number of Inventions: 1

(4 pages total)

(54) Title of the Invention: Kana and Kanji Converter

(21) Application No.: 59-155977

(22) Application Date: July 26, 1984

(72) Inventor: Yasuyuki Numata
c/o Ricoh Co., Ltd.
1-3-6 Nakagome
Ota-ku, Tokyo

(71) Applicant: Ricoh Co., Ltd.
1-3-6 Nakagome
Ota-ku, Tokyo

(74) Representative: Masatoshi Isonuma, Patent Attorney

SPECIFICATIONS

1. Title of the Invention

Kana and Kanji Converter

2. Claim

(1) Kana and kanji converter so characterized that in a kana and kanji converter that has a best candidate selector that selects the best candidate word from among a group of candidate words extracted by searching a dictionary, it is fitted with a means that evaluates whether or not the best candidate selected by the above-mentioned best candidate selector has a high probability of reappearing, a memory means that stores the best candidate evaluated by said means as having the highest probability of reappearing together with its distinguishing information, a means that substitutes the best candidate loaded first with the best candidate last selected when the best candidates loaded in said memory means exceed a set number, and a display means that displays the content of the above-mentioned memory means, and is constructed such that by inputting the distinguishing information of a best candidate displayed by said display means, the best candidate corresponding to said distinguishing information is outputted as the conversion result.

3. Detailed Description of the Invention

Technical Field

This invention pertains to a kana and kanji converter. More particularly, this invention pertains to a kana and kanji converter that can be applied to devices such as Japanese language word processors or electronic calculator systems.

Prior Art

Kana and kanji converters convert text inputted as phonetic text (hiragana, katakana, or Latin alphabet) to an appropriate mixture of kanji and kana. Therefore, these have a word dictionary for converting to kana and kanji installed. This word dictionary is searched by separating words from the inputted kana text string to form search text strings, then matching the search text strings to text strings found in the word dictionary. However, because Japanese has a complicated grammar and many homophones that differ in meaning, many candidate words are extracted by searching the dictionary.

To select one from among these many candidate words as the conversion result, each of the candidate words extracted is evaluated as follows: Factors such as the length, use frequency, and connection probability of the candidate word are evaluated as parameters, and the candidate word given the highest number of evaluation points (the best candidate word) is outputted as the conversion result.

It is a characteristic of texts that they repeat the same words. However, because input methods by prior art do not

consider this, when the same words appear frequently in a text, the same word reading must be inputted each time, and this poor input efficiency increases the physical and mental exhaustion of the operator. Moreover, the difficulty described above is increased still further in devices that require pressing a conversion key to convert to kana and kanji each time the operator finishes inputting a text string corresponding to kanji.

In addition, as far as a kana and kanji converter is concerned, repeating a complicated conversion to kana and kanji each time the same word appears is undesirable in terms of conversion speed.

Purpose

The purpose of this invention is to solve problems in prior art such as described above by offering a kana and kanji converter that can avoid repeated input of the same word and increase text input efficiency.

Structure

To achieve the purpose given above, the structure of this invention is so characterized that in a kana and kanji converter that has a best candidate selector that selects the best candidate word from among a group of candidate words extracted by searching a dictionary, it is fitted with a means that evaluates whether or not the best candidate selected by the above-mentioned best candidate selector have a high probability of reappearing, a memory means that stores the best candidate evaluated by said means as having the highest probability of reappearing together with its distinguishing information, a means that substitutes the best candidate loaded first with the best candidate last selected when the best candidates loaded in said memory means exceed a set number, and a display means that displays the content of the above-mentioned memory means, and is constructed such that by inputting the distinguishing information of a best candidate displayed by said display means, the best candidate corresponding to said distinguishing information is outputted as the conversion result.

Below, the structure of this invention is explained in greater detail by a working example.

Figure 1 is a schematic block diagram of a kana and kanji converter according to a working example of this invention.

In Figure 1, 1 is a text inputter, 2 is a dictionary searcher, 3 is a word dictionary, 4 is a word connection detector, 5 is a word connection detection chart, 6 is an evaluator, 7 is a best candidate selector, 8 is a best candidate memory, 9 is a distinguishing information inputter, 10 is a condition evaluator, 11 is a word holder, 12 is a word substitution circuit, and 13 is a display.

Dictionary searcher 2 searches word dictionary 3 for the input text string inputted by text inputter 1 and extracts

gets context from previous correspondence - or from keyword content in memory

MUST FINISH
INPUT (K-K CONSENSI
OR R-K
CONV.)

candidate words. Word connection detector 4 uses word connection detection chart 5 to detect whether or not the candidate words extracted can be connected to the word converted just before. Evaluator 6 evaluates the candidate words that can be connected using parameters such as length, use frequency, and seriousness of the connection. Best candidate selector 7 selects the candidate word that has the highest evaluation and loads this into best candidate memory 8.

Figure 2 is a flow chart that shows the operation of storing the word with the highest probability of reappearing together with its distinguishing information.

First, best candidate selector 7 selects the best candidate (201) and sends it to condition evaluator 10.

Condition evaluator 10 evaluates whether or not the best candidate sent from best candidate selector 7 satisfies the condition loaded in word holder 11 (202). For this evaluation, different conditions can be set depending on the field of the text. Generally, it is most effective to make the condition whether or not the word is a speech part such as a noun.

When the condition is not satisfied, if text input has not ended, condition evaluator 10 requests the next best candidate from best candidate selector 7 (202, 210, 201). When the condition is satisfied, it is checked whether or not the same best candidate is loaded already in word holder 11 (203). When it is loaded already, the record sequence is modified, and if text input has not yet ended, the next best candidate is requested (203, 204, 210, 201).

When the best word is not recorded in word holder 11, it is evaluated whether or not there is still space in word holder 11 (205). When there is space, said best candidate is recorded in the empty space and the record sequence is set as distinguishing information (205, 206, 209). When there is no space, word substitution circuit 12 erases the word recorded the longest time ago (205, 207). The best candidate is recorded in this space, the record sequence is set, and if input has not yet ended, the following best candidate is requested (208, 209, 210, 201).

For example, it is assumed that the following text is inputted:

Sample Text:

Greetings. I am glad that you continue to enjoy good health this midsummer. I have benefitted from your special kindness and favor in the past, and I wish to thank you from my heart.

Today, I am notifying you that at a meeting of the Board of Directors the other day and at the 26th General Accounting Meeting, our company decided the following:

1 The Osaka branch will close on July 31, 19●●, and its assets and liabilities will be taken over by the new company of Morita Commercial Co., Ltd.

2 Director and Osaka Branch Chief Kozo Morita of Morita ● Industries (Ltd.) will be appointed

Representative Director of the new company of Morita Commercial Co., Ltd., and take over its business.

Moreover, in this working example, it is assumed that the condition for recording in word holder 11b is whether or not a word is a kanji noun with two or more kanji such as shown underlined in the sample text given above. In addition, it is assumed that word holder 11 can record twenty words.

In the case of the sample text given above, words that satisfy this condition such as "greetings," "midsummer," etc. are recorded in sequence in word holder 11 and displayed by display 13.

Figure 3 shows the display content of display 13 when the memory capacity of word holder 11 is first filled to capacity.

When "close" is inputted with word holder 11 full as shown in Figure 3, "1. greetings" is replaced by "1. close" and the display content becomes as shown in Figure 4. Following this, when "assets," "liabilities," and "company" are inputted, the display content becomes as shown in Figure 5.

Next, when "Morita" is to be inputted, because there is already an entry "17. Morita" in slot 17 of the words displayed, this may be inputted simply by inputting this distinguishing information "17" by distinguishing information inputter 9. This avoids repeated input of the same word, and can raise text input efficiency. In addition, when inputted by distinguishing information, there is no need to perform a dictionary search, word connection detection, and evaluation, and conversion speed can be improved.

Effects

As explained above, the kana and kanji converter according to this invention avoids repeated input of the same words and can raise text input efficiency.

4. Brief Explanation of the Figures

Figure 1 is a schematic block diagram of a kana and kanji converter according to a working example of this invention. Figure 2 is a flow chart that shows the operation of storing a word in the word holder in Figure 1. Figure 3, Figure 4, and Figure 5 are diagrams that show the content of the word holder displayed by the display.

7: best candidate selector, 8: best candidate memory, 9: distinguishing information inputter, 10: condition evaluator, 11: word holder, 12: word substitution circuit, 13: display

Patent Applicant: Ricoh Co., Ltd.

Representative: Masatoshi Isonuma

Figure 1

1 ... text inputter, 2 ... dictionary searcher, 3 ... word dictionary, 4 ... word connection detector, 5 ... word connection detection chart, 6 ... evaluator, 7 ... best candidate selector, 8 ... best candidate memory, 9 ... distinguishing information inputter, 10 ... condition evaluator, 11 ... word holder, 12 ... word substitution circuit, 13 ... display

Figure 2

201 ... select best candidate, 202 ... does this satisfy condition for being held in the word holder?, 203 ... is the same word loaded in the word holder?, 204 ... change record sequence, 206 ... record in empty space, 207 ... erase the word recorded the longest time ago, 208 ... record in erased space, 209 ... set record sequence, 210 ... has text input ended?

Figure 3

1. greetings	2. midsummer	3. health	4. in the past	5. special
6. kindness	7. favor	8. our company	9. the other day	10. Board of Directors
11. assets	12. liabilities	13. following	14. decided	15. notify
16. 19 hundred	17. Morita	18. industries	19. Osaka	20. branch

Figure 4

1. close	2. midsummer	3. health	4. in the past	5. special
6. kindness	7. favor	8. our company	9. the other day	10. Board of Directors
11. assets	12. liabilities	13. following	14. decided	15. notify
16. 19 hundred	17. Morita	18. industries	19. Osaka	20. branch

Figure 5

1. close	2. assets	3. liabilities	4. company	5. special
6. kindness	7. favor	8. our company	9. other day	10. Board of Directors
11. assets	12. liabilities	13. following	14. decided	15. notify
16. 19 hundred	17. Morita	18. industries	19. Osaka	20. branch

⑨ 日本国特許庁(JP)

⑩ 特許出願公開

⑫ 公開特許公報(A)

昭61-33570

⑪ Int. Cl.

識別記号

庁内整理番号

⑬ 公開 昭和61年(1986)2月17日

G 06 F 15/20

A-7010-5B

審査請求 未請求 発明の数 1 (全4頁)

⑭ 発明の名称 カナ漢字変換処理装置

⑮ 特 願 昭59-155977

⑯ 出 願 昭59(1984)7月26日

⑰ 発 明 者 沼 田 泰 之 東京都大田区中馬込1丁目3番6号 株式会社リコー内
⑱ 出 願 人 株 式 会 社 リ コ ー 東京都大田区中馬込1丁目3番6号
⑲ 代 理 人 弁 理 士 磯 村 雅 俊

沼田 泰之

1. 発明の名称 カナ漢字変換処理装置

2. 特許請求の範囲

(1) 辞書検索により抽出された候補単語群の中から最適な候補単語を選出する最適候補選出部を有するカナ漢字変換処理装置において、前記最適候補選出部により選出された最適候補の再出蓋然性が高いか否かを判定する手段と、該手段により再出蓋然性が高いものであると判定された最適候補をその識別情報とともに記憶する記憶手段と、該記憶手段に格納された最適候補が一定数を越える時、最初に格納した最適候補と最後に選出された最適候補とを入れ替える手段と、前記記憶手段の内容を表示する表示手段を設け、該表示手段に表示された最適候補の識別情報を入力することにより、該識別情報に対応する最適候補を変換結果として出力するよう構成したことを特徴とするカナ漢字変換処理装置。

3. 発明の詳細な説明

技術分野

本発明はカナ漢字変換処理装置に関し、より詳細には日本語文書作成装置、電子計算機システム等に適用し得るカナ漢字変換処理装置における文字列入力方式に関する。

従来技術

カナ漢字変換処理装置では、表音文字(ひらがな、カタカナ、ローマ文字)で入力された文章を適切な漢字カナ混じり文に変換するため、カナ漢字変換用の単語辞書を設けている。この単語辞書の検索は、入力されたカナ文字列から単語を切出して被検索文字列とし、被検索文字列と単語辞書中の見出し文字列とのマッチングを行うことにより行う。しかし、日本語は文法が複雑で、かつ同音異義語が多数存在するため、辞書検索により複数の候補単語が抽出される。

この複数の候補単語の中から1つを選択し変換結果とするために、抽出された候補単語のそれぞれに対し、次のような評価を行っている。すなわち、候補単語を読み長、使用頻度、接続確率等を

パラメータとして評価し、評価値の最も高い候補単語（最適候補単語）を交換結果として出力する。

ところで、文章には、同一の単語が繰り返し出現するという特徴がある。しかし、従来の入力方式においては、このことを考慮していないため、文章中に同一の単語が頻りに現われる場合においても、その都度、同一の単語の読みを入力する必要があり入力効率が悪い。オペレータの肉体的、精神的疲労が増大する。また、漢字相当の文字列を入力し終るたびにカナ漢字交換のための交換キーを押下する装置の場合は、さらに上記弊害が大きくなる。

一方、カナ漢字交換処理装置にとっても、同一の単語が出現するたびに複雑なカナ漢字交換処理を繰り返すことは、交換処理速度の点から不具合である。

目 的

本発明の目的は、上記のような従来技術の問題点を解決するため、同一単語の繰り返し入力を回避し、文字入力の効率化を図り得るカナ漢字交換

処理装置を提供することにある。

構 成

以上の目的を達成するため、本発明の構成は、辞書検索により抽出された候補単語群の中から最適な候補単語を選出する最適候補選出部を有するカナ漢字交換処理装置において、前記最適候補選出部により選出された最適候補の再出蓋然性が高いか否かを判定する手段と、該手段により再出蓋然性が高いものであると判定された最適候補をその識別情報とともに記憶する記憶手段と、該記憶手段に格納された最適候補が一定数を越える時、最初に格納した最適候補と最後に選出された最適候補とを入れ替える手段と、前記記憶手段の内容を表示する表示手段を設け、該表示手段に表示された最適候補の識別情報を入力することにより、該識別情報に対応する最適候補を交換結果として出力するよう構成したことに特徴がある。

以下、本発明の構成を一実施例により詳細に説明する。

第1図は、本発明の一実施例によるカナ漢字交

換処理装置の概略ブロック図である。

第1図において、1は文字入力部、2は辞書検索部、3は単語辞書、4は接続可否検定部、5は接続可否検定表、6は評価部、7は最適候補選出部、8は最適候補記憶部、9は識別情報入力部、10は条件判定部、11は単語保持部、12は単語入替回路、13は表示部である。

辞書検索部2は、文字入力部1により入力された入力文字列に対して、単語辞書3を検索し、候補単語を抽出する。接続可否検定部5は抽出された候補単語に対し、接続可否検定表5を用いて直前の交換済単語との接続の可否を検定する。評価部6は、接続可能な候補単語を読み長、出現頻度、接続の重み等をパラメータとして評価する。最適候補選出部7は、最も評価の高い候補単語を選出し、それを最適候補単語記憶部8に格納する。

第2図は、再出する蓋然性の高い単語を、その識別情報とともに記憶する動作を示すフローチャートである。

まず、最適候補選出部7は最適候補を選出し

(201)、条件判定部10に送出する。

条件判定部10は、最適候補選出部7から送出された最適候補が単語保持部11に格納する条件を満足するか否かを決定する(202)。この場合、文章の分野により異なる条件を設定することも可能である。一般的には、名詞等の品詞であるか否かを条件にした方が効果的である。

条件を満足しない場合は、条件判定部10は、文字入力部1が終了していなければ、最適候補選出部7に対して次の最適候補を要求する(202, 210, 201)。条件を満足する場合は、すでに単語保持部11に同一の最適候補が格納されているか否かを調べる(203)。すでに格納されている場合は登録順位を変更し、文字入力部1が終了であれば、次の最適候補を要求する(203, 204, 210, 201)。

単語保持部11に登録されていない最単語の場合は、単語保持部11にまだ空きがあるか否かを判定する(205)。空きがある場合は、空きの部分に当該最適候補を登録し、識別情報としての

登録順位をセットする(205, 206, 209)。空きがない場合は、単語入替回路12は、最も以前に登録された単語を抹消し(205, 207)、その部分に最適候補を登録し、登録順位をセットして、入力未終了であれば後続の最適候補を要求する(208, 209, 210, 201)。

例えば、次のような例文を入力するものとする。

例文：

拝啓 盛夏の候、ますますご清栄のこととお喜び申し上げます。平素は格別のご恩恵、ご高配を隔り、心からありがたく厚くお礼申し上げます。

さて、当社では、先般行われました取締役会および第26回決算総会において下記のとおり決定いたしましたのでご通知申し上げます。

1 来る昭和〇〇年7月31日をもって森田〇〇工業(株)大阪支店を閉鎖し、債権債務は新会社森田産業株式会社が継承する。

2 森田〇〇工業(株)の取締役大阪支店長森田広三は、新会社森田産業株式会社代表取締役に就任し、事業を継承する。

なお、本実施例では、上記例文中、下線で示したような漢字2字以上の漢語の名称であるか否かが単語保持部11に登録するための条件であるものとする。また、単語保持部11は20個の単語を登録し得るものとする。

上記例文の場合、「拝啓」、「盛夏」…等、条件を満たす単語は単語保持部11に順次登録され、表示部13に表示される。

第3図は、単語保持部11の記憶容量一杯まで記憶した最初の状態における表示部13の表示内容を示す。

単語保持部11が第3図の状態で、「閉鎖」を入力すると、「1. 拝啓」が「1. 閉鎖」に代わり、第4図に示したような表示内容となる。続いて「債権」、「債務」、「会社」を入力すると、第5図に示したような表示内容となる。

続く「森田」を入力するにあたっては、既に表示単語の中の17番に「17. 森田」があるので、この17という簡単な識別情報を識別情報入力部9から入力すれば良く、同一の単語の繰り返し入

力を回避し、文字入力の効率化を図ることができる。また、識別情報による入力の場合は、辞書検索、接続可否検索、評価を行う必要がなく変換処理速度を向上することができる。

効果

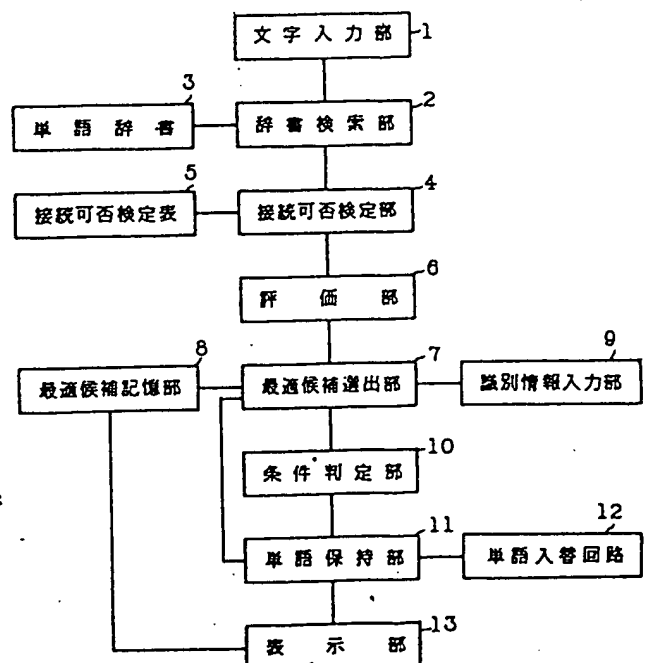
以上説明したように、本発明によるカナ漢字変換処理装置によれば、同一単語の繰り返し入力を回避し、文字入力の効率化を図ることができる。

4. 図面の簡単な説明

第1図は本発明の一実施例によるカナ漢字変換処理装置の概略ブロック図、第2図は第1図の単語保持部への単語格納動作を示すフローチャート、第3図、第4図、第5図は単語保持部の内容を表示部に表示した図である。

7：最適候補選出部、8：最適候補記憶部、9：識別情報入力部、10：条件判定部、11：単語保持部、12：単語入替回路、13：表示部。

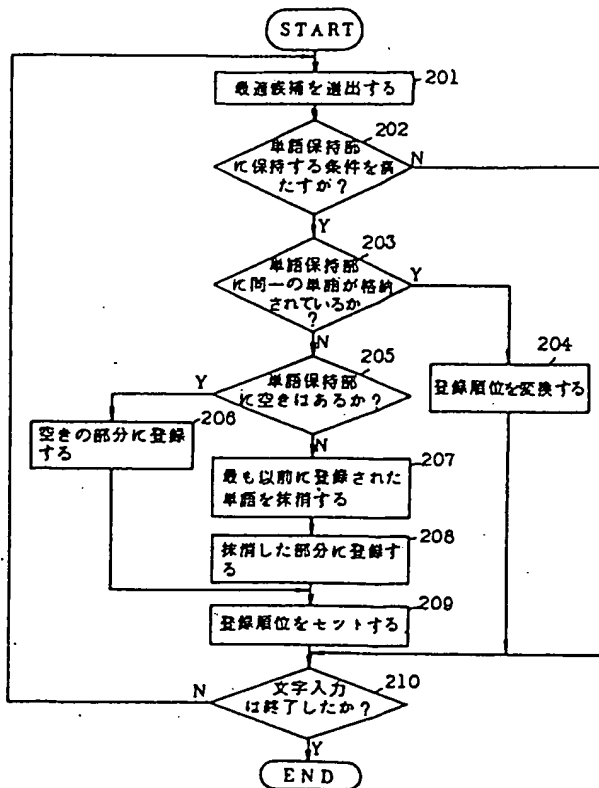
第 1 図



特許出願人 株式会社 リ コ

代理人 井理士 磯村 隆

第 2 図



第 3 図

1. 辞書	2. 盛夏	3. 清栄	4. 平素	5. 格別
6. 懸情	7. 高配	8. 当社	9. 先般	10. 取締役
11. 決算	12. 総会	13. 下記	14. 決定	15. 通知
16. 昭和	17. 森田	18. 工業	19. 大阪	20. 支店

第 4 図

1. 閉鎖	2. 盛夏	3. 清栄	4. 平素	5. 格別
6. 懸情	7. 高配	8. 当社	9. 先般	10. 取締役
11. 決算	12. 総会	13. 下記	14. 決定	15. 通知
16. 昭和	17. 森田	18. 工業	19. 大阪	20. 支店

第 5 図

1. 閉鎖	2. 債券	3. 債務	4. 会社	5. 格別
6. 懸情	7. 高配	8. 当社	9. 先般	10. 取締役
11. 決算	12. 総会	13. 下記	14. 決定	15. 通知
16. 昭和	17. 森田	18. 工業	19. 大阪	20. 支店

265